

KOREAN INTELLECTUAL PROPERTY OFFICE

KOREAN PATENT ABSTRACTS

(11)Publication number: 1020020061443 A
 (43)Date of publication of application: 24.07.2002

(21)Application number: 1020010002853

(71)Applicant: TOBESOFT CO., LTD.

(22)Date of filing: 18.01.2001

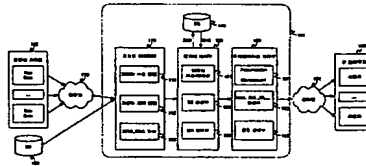
(72)Inventor: KIM, HYEONG GON

(51)Int. Cl. G06F 17/30

(54) METHOD AND SYSTEM FOR COLLECTING, PROCESSING, AND DISPLAYING INFORMATION USING
 COMPUTER COMMUNICATION NETWORK

(57) Abstract:

PURPOSE: A method and a system for collecting, processing, and displaying information are provided to search information desired by a user from web documents via the Internet, and display the searched information as a desired form.



CONSTITUTION: A system(100) for collecting, analyzing and processing information is connected to various Internet sites(150) and a user web browser(180) via the Internet(170). The system(100) receives a search word of information from the user via the user web browser(180).

The system(100) collects information from the Internet sites(150), and analyzes and processes collected data before providing it to the user. The system(100) is connected to an external database(160) directly or via the Internet(170), so that the system(100) collects information from the external database(160). The system(100) includes an intelligent-type agent(110), a data process part(120) and a presentation process part(130). The system(100) further includes a database(140) for storing collected, processed information.

&copy; KIPO 2003

Legal Status

Date of final disposal of an application (20040921)

Date of registration (00000000)

Date of opposition against the grant of a patent (00000000)

BEST AVAILABLE COPY

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(51) Int. Cl. 7
G06F 17/30

(11) 공개번호 특2002-0061443
(43) 공개일자 2002년07월24일

(21) 출원번호 10-2001-0002853
(22) 출원일자 2001년01월18일

(71) 출원인 (주)투비소프트
서울특별시 강남구 대치동 997-10 메디슨벤처타워 별관 504호
(72) 발명자 김형곤
서울특별시송파구방이동89올림픽선수촌아파트204-702
(74) 대리인 김삼수
송정은

심사청구 : 있음

(54) 컴퓨터 통신망을 이용한 정보의 수집, 가공 및 표시방법과 그 시스템

요약

사용자가 원하는 정보를 찾기 위한 검색어를 입력하고 검색 대상을 지정하면, 입력된 검색어를 분석한 후 데이터 수집 엔진을 이용하여 원하는 정보가 포함된 웹 문서, 기존 데이터베이스 등을 검색하여 원하는 정보가 포함된 자료를 수집한다. 수집된 자료로부터 자연어 처리 엔진 또는 HTML/XML 파서 등을 이용하여 원하는 정보를 추출한 후, 추출된 정보를 표준화/정형화하여 데이터베이스 형태로 저장한다. 필요에 따라 미리 생성되어 저장되어 있는 데이터베이스에 포함된 자료를 병합하여 검색된 자료와 함께 시각적인 형태로 사용자에게 표시하거나, 재가공할 수 있는 형태로 데이터를 가공하여 사용자에게 제공한다. 표시 형태는 사용자의 요구에 따라 여러가지로 달라질 수 있는데, 그래프나 표 등을 포함하는 웹 문서나 보고서 형태가 될 수 있으며, 저장된 데이터를 다시 문장으로 생성하여 보고서에 포함되도록 할 수도 있다. 또한, 데이터베이스에 저장된 데이터는 재가공할 수 있는 스프레드시트 파일로 포스팅될 수 있다.

대표도
도 1

색인어
문서 검색, 정보 수집, 정보 가공, 지능형 에이전트, 데이터 표준화/정형화, 보고서 작성

명세서

도면의 간단한 설명

도 1은 본 발명의 컴퓨터 통신망을 이용한 정보 수집, 분석 및 가공 방법을 구현하는 전체 시스템의 구성을 나타내는 개략도이다.

도 2는 본 발명의 컴퓨터 통신망을 이용한 정보 수집, 분석 및 가공 방법을 나타내는 흐름도이다.

도 3은 본 발명의 컴퓨터 통신망을 이용한 정보 수집, 분석 및 가공 방법에서 원하는 정보를 포함하는 웹 문서를 검색하여 웹 문서로부터 정보를 추출하는 과정을 시각적으로 표시한 도면이다.

도 4는 본 발명의 컴퓨터 통신망을 이용한 정보 수집, 분석 및 가공 방법에서 수집한 정보를 정형화하는 과정을 나타내는 도면이다.

도 5는 본 발명의 컴퓨터 통신망을 이용한 정보 수집, 분석 및 가공 방법에서 수집한 정보와 병합하기 위한 기존의 정보를 포함하고 있는 데이터베이스의 예를 나타내는 도면이다.

도 6은 본 발명의 컴퓨터 통신망을 이용한 정보 수집, 분석 및 가공 방법에서 수집한 정보를 표와 그래프 형태로 표시해주는 과정을 나타내는 도면이다.

도 7은 본 발명의 컴퓨터 통신망을 이용한 정보 수집, 분석 및 가공 방법에서 수집한 정보를 이용하여 전략보고서를 생성하는 과정을 나타내는 도면이다.

도 8은 본 발명의 컴퓨터 통신망을 이용한 정보 수집, 분석 및 가공 방법에서 수집한 정보를 재처리가능한 스프레드시트 파일로 포스팅하는 과정을 나타내는 도면이다.

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 컴퓨터 통신망을 이용하여 정보를 수집, 분석 및 가공하는 방법 및 그 시스템에 관한 것으로서, 구체적으로는 사용자가 필요한 정보에 대한 검색어를 입력하면 검색 엔진을 이용하여 인터넷의 웹 문서 상의 각종 관련 정보를 수집하여, 수집한 문서내에서 필요한 데이터만을 추출하고 데이터베이스화하여 이를 재가공하여 사용자에게 제공하는 방법 및 그 시스템에 관한 것이다.

인터넷은 정보의 바다라고 불리고 있는 것처럼, 세계 곳곳에 존재하는 수많은 정보를 인터넷을 통해 얻을 수 있다. 그러나, 정보의 양이 지나치게 방대하고, 정보의 표시 형식 등이 체계화되어 있지 않으므로 모든 정보를 일일이 사용자가 검색하여 원하는 정보를 찾는다는 것은 불가능하다. 이와 같은 점을 고려하여 사용자가 원하는 정보를 찾는 과정을 도와주기 위한 많은 방법들이 사용되고 있다.

대표적인 것으로는 야후나 엠파스와 같은 검색 엔진을 들 수 있는데, 이러한 검색 엔진들은 사용자가 찾기를 원하는 정보에 관한 검색어를 입력하면, 인터넷의 웹 문서들을 검색하여 관련된 웹 문서 또는 사이트를 찾아 해당 웹 문서 또는 사이트로의 링크를 제공하고 미리보기 기능 등을 제공한다.

그밖에도 원하는 정보가 포함되어 있는 다수의 사이트로부터 특정 영역들을 조합하여 사용자의 브라우저를 통해 동시에 볼 수 있도록 해주는 맞춤형 브라우징(personalized browsing) 서비스 등이 제공되고 있다.

그러나, 이와 같은 방법들은 모두 검색 대상 정보를 표시해주거나, 검색 대상 정보에 대한 링크를 제공할 뿐이어서, 사용자는 검색된 정보들을 모두 일일이 확인하여야 하며, 더 나아가 검색된 정보를 재이용하기 위해서는 사용자가 수집된 정보를 다시 데이터베이스에 입력하고 보고서를 작성하는 등의 많은 작업을 필요로 한다.

한편, 이와 같은 문제점을 해결하기 위한 방법으로 가격 비교 엔진과 같은 지능형 에이전트(Intelligent agent)가 사용되고 있다. 이러한 지능형 에이전트는 학습된 규칙(rule)에 따라 원하는 정보를 수집하여 수집된 특정 정보에 대한 비교 정보를 제시해줄 수 있다. 그러나, 지능형 에이전트에 의한 정보 수집 및 비교 방식 역시 그 대상 정보가 특정한 정보에 한정되어 있고(예를 들면, 판매가격), 한정된 정보에 대한 비교정보를 나열하는 것에 불과하고, 비교정보에 대해서도 단순히 표시하는 것일 뿐 정보를 재가공할 수 있는 형태로 사용자에게 제공하는 것은 아니다.

따라서, 단순히 원하는 정보를 표시해주는 것에 그치지 않고, 원하는 정보를 찾아 분석하고 가공하여 사용자에게 시각적인 형태로 표시해주거나, 재가공이 용이한 형태로 데이터베이스화해 준다면 정보 검색에 드는 시간과 비용을 크게 줄일 수 있고, 개별화된 정보(컨텐츠) 서비스를 위한 기반 솔루션으로도 활용할 수 있는 등 매우 큰 효과를 얻을 수 있을 것이다.

발명이 이루고자 하는 기술적 과제

본 발명은 이러한 점을 감안하여 이루어진 것으로서, 인터넷을 통해 웹 문서 등으로부터 원하는 정보를 찾아 이를 사용자가 원하는 형태로 표시해줄 수 있는 정보 수집, 가공 및 표시 방법 및 그 시스템을 제공하는 것을 그 목적으로 한다.

본 발명의 다른 목적은 인터넷 상에서 검색하여 수집한 문서로부터 원하는 정보를 추출하여 이를 재가공할 수 있는 형태로 사용자에게 제공할 수 있는 정보 수집, 가공 및 표시 방법 및 그 시스템을 제공하는 것이다.

발명의 구성 및 작용

이와 같은 목적을 달성하기 위하여 본 발명에서는, 사용자가 원하는 정보를 찾기 위한 검색어를 입력하고 검색 대상을 지정하면, 입력된 검색어를 분석한 후 데이터 수집 엔진을 이용하여 원하는 정보가 포함된 웹 문서, 기존 데이터베이스 등을 검색하여 원하는 정보가 포함된 자료를 수집한다. 수집된 자료로부터 자연어 처리 엔진 또는 HTML/XML 파서 등을 이용하여 원하는 정보를 추출한 후, 추출된 정보를 표준화/정형화하여 데이터베이스 형태로 저장한다. 필요에 따라 미리 생성되어 저장되어 있는 데이터베이스에 포함된 자료를 병합하여 검색된 자료와 함께 시각적인 형태로 사용자에게 표시하거나, 스프레드시트 파일과 같은 재가공할 수 있는 형태로 데이터를 가공하여 사용자에게 제공한다. 표시 형태는 사용자의 요구에 따라 정해줄 수 있으며, 그래프나 표 등을 포함하는 웹 문서나 보고서 형태로 될 수 있고, 저장된 데이터를 다시 문장으로 생성하여 보고서에 포함되도록 할 수도 있다.

즉, 본 발명에 따른 컴퓨터 통신망을 이용한 정보 수집, 가공 및 표시 방법은, 컴퓨터 통신망을 통하여 사용자로 부터 검색어를 입력받는 단계, 상기 검색어를 이용하여 상기 검색어가 지시하는 정보가 포함된 웹 문서 또는 데이터베이스를 검색하는 단계, 검색된 웹 문서 또는 데이터베이스로부터 상기 검색어가 지시하는 정보를 추출하는 단계, 추출된 정보를 표준화/정형화하여 데이터베이스의 형태로 저장하는 단계, 데이터베이스의 형태로 저장된 정보를 표시할 형태를 컴퓨터 통신망을 통하여 사용자로 부터 입력받는 단계, 데이터베이스의 형태로 저장된 정보를 사용자로 부터 입력된 표시형태에 따라 컴퓨터 통신망을 통하여 사용자에게 제공하는 단계를 포함하여 이루어진다.

여기에서, 상기 정보를 표시할 형태를 사용자로 부터 입력받는 단계에서는, 상기 정보를 포함하는 표, 상기 정보를 나타내는 그래프, 상기 정보를 포함하는 문장, 상기 정보를 포함하는 재가공 가능한 스프레드시트용 파일 중 적어도 하나를 포함하도록 사용자가 표시 형태를 직접 디자인할 수 있도록 하거나, 상기 정보를 포함하는 다양한 표시 형태를 갖는 템플릿을 컴퓨터 통신망을 통하여 사용자에게 제공하고, 사용자가 상기 템플릿 중에서 원하는 것을 선택하도록 할 수 있다.

한편, 본 발명에 따른 컴퓨터 통신망을 이용한 정보 수집, 가공 및 표시 시스템은, 컴퓨터 통신망을 통하여 사용자로부터 입력된 검색어를 입력받는 입력부, 입력부로부터 검색어를 제공받아 상기 검색어가 지시하는 정보가 포함된 웹 문서 또는 데이터베이스를 검색하는 데이터 수집 엔진, 데이터 수집 엔진으로부터 검색된 웹 문서 또는 데이터베이스로부터 상기 검색어가 지시하는 정보를 추출하는 정보 추출부, 정보 추출부로부터 추출된 정보를 표준화/정형화하여 저장하는 데이터 처리부, 표준화/정형화된 정보를 저장하는 데이터베이스, 데이터베이스에 저장된 정보를 컴퓨터 통신망을 통하여 사용자에게 제공하는 표시부를 포함한다.

여기에서, 정보 추출부는 검색된 상기 웹 문서로부터 HTML/XML 태그를 제거하고 정보가 포함된 부분만을 추출할 수 있는 HTML/XML 파서와 검색된 상기 웹 문서에 포함된 문장으로부터 상기 검색어가 지시하는 정보를 추출할 수 있는 자연어 처리부를 포함하는 것이 바람직하다.

이제 본 발명의 바람직한 실시예에 대하여 도면을 참고로 하여 상세히 설명한다.

도 1은 본 발명의 컴퓨터 통신망을 이용한 정보 수집, 분석 및 가공 방법을 구현하는 전체 시스템의 구성을 나타내는 개략도이다.

본 발명의 정보 수집, 분석 및 가공 시스템(100)은 인터넷(170)을 통하여 다양한 인터넷 사이트(150)들과 사용자의 웹 브라우저(180)에 연결될 수 있으며, 사용자의 웹 브라우저(180)를 통해 사용자로부터 수집할 정보에 대한 검색어를 입력받아, 인터넷 사이트(150)로부터 정보를 수집하여, 데이터를 분석 및 가공한 후 사용자에게 제공한다. 또한, 본 발명의 시스템(100)은 미리 구축되어 있는 외부의 데이터베이스(160)들과 직접 또는 인터넷을 통해 연결될 수 있으며, 이와 같은 외부의 데이터베이스(160)들로부터 데이터를 수집할 수도 있다. 인터넷 사이트(150)나 외부 데이터베이스(160)와 같은 정보원으로부터 수집된 정보는 본 발명의 시스템(100) 내에서 통합적으로 처리된다.

본 발명의 시스템(100)은 크게 지능형 에이전트(110)와 데이터 처리부(120) 및 프리젠테이션 처리부(130)로 이루어져 있으며, 수집하여 처리한 정보를 저장하기 위한 데이터베이스(140)를 구비하고 있다.

지능형 에이전트(110)는 데이터 수집 및 분석에 관계한다. 지능형 에이전트(110)는 다시 인터넷 사이트들을 통해 원하는 데이터를 검색하여 수집하는 데이터 수집 엔진(111), 문장 등으로 이루어진 데이터를 처리하기 위한 자연어 처리 엔진(112), HTML(HyperText Markup Language) 문서 또는 XML(Extensible Markup Language) 문서로 된 데이터로부터 원하는 데이터를 추출하기 위한 HTML/XML 파서(113)로 구성되어 있으며, 지능형 에이전트(110)에 의해 수집되고 처리된 데이터는 데이터 처리부(120)로 전달된다.

데이터 처리부(120)는 지능형 에이전트(110)로부터 전달받은 데이터를 일정한 형태로 표준화/정형화하기 위한 데이터 표준/정형화부(121)와 이를 이용하여 데이터베이스를 생성하는 데이터베이스 생성부(122), 생성된 데이터베이스를 제어하기 위한 데이터베이스 제어부(123) 등으로 구성된다. 데이터베이스 생성부(122)에 의해 생성된 데이터베이스(140)는 사용자의 필요에 따라 언제든지 불러내어 사용될 수 있다.

데이터 처리부(120)에 의해 표준화/정형화되어 데이터베이스(140)에 저장된 데이터를 사용자에게 필요한 형태로 표시하는 것은 프리젠테이션 처리부(130)에 의해 처리된다. 프리젠테이션 처리부(130)는 다시 프리젠테이션 컴포넌트(131), html, XML, XSL 생성부(132), 문장 생성부(133) 등으로 구성되어, 사용자의 요청에 따라 원하는 형태의 프리젠테이션을 생성하고 이를 다시 사용자에게 제공한다.

이제, 이와 같은 시스템을 이용하여 구현되는 본 발명의 컴퓨터 통신망을 이용한 정보 수집, 분석 및 가공 방법에 대해 설명한다.

도 2는 본 발명의 컴퓨터 통신망을 이용한 정보 수집, 분석 및 가공 방법을 나타내는 흐름도이다.

먼저, 사용자는 원하는 정보를 찾기 위해 검색어를 입력하고, 검색 대상을 선정한다(S210). 검색어는 단어나 문장으로 입력할 수 있으며, 검색대상은 웹 문서, 기존의 지식관리 시스템(KMS; Knowledge Management System)이나 기타 데이터베이스를 복수로 지정할 수 있다. 웹 문서의 경우 특정 사이트를 지정하여 검색할 수도 있다. 사용자가 검색 대상을 지정하지 않는 경우, 디폴트로 검색할 검색 대상을 미리 지정해둘 수도 있다. 예를 들면, 사용자는 "중국의 인터넷 광고시장 규모는?" 과 같은 질문을 검색어로 입력할 수 있다.

사용자가 검색어와 검색 대상을 입력하면, 이를 이용해 검색 대상으로 지정된 웹 문서, 콘텐츠 데이터베이스, 기존 데이터베이스 등을 검색하여 필요한 정보를 수집한다(S220). 이러한 정보 수집 과정은 지능형 에이전트(110) 내의 데이터 수집 엔진(111)을 통해 하게 되는데, 이는 운영자가 웹 문서를 반복적으로 검색하여 규칙(rule)을 정의하고 이를 학습시키는 방식으로 수행된다. 이러한 방식은 종래의 검색 엔진(로봇) 등에 이미 구현되어 있으며, 이와 같은 알려진 방법을 사용하여 데이터를 수집할 수 있다. 예를 들면, 가격비교 사이트의 경우, 특정 쇼핑물의 가격정보가 있는 링크들을 서버가 추적하고 이 규칙을 검색 에이전트에 학습시켜, 검색 에이전트가 규칙을 따라 링크를 추적하여 가격 정보를 가져오도록 하고 있는데, 본 발명의 데이터 수집 단계에서도 이와 같은 방식을 통해 데이터를 수집할 수 있다.

다음 단계는 검색된 다양한 형태의 정보들로부터 원하는 정보를 분석하고 추출해내는 것이다(S230). 이 단계에서는 자연어 처리 엔진(112)이나 HTML/XML 파서(113) 등이 이용된다. 즉, 필요한 경우 웹 문서에 HTML/XML 파싱(parsing) 기술을 적용하여 HTML/XML 태그를 제외하고 필요한 데이터만을 추출해내고, 검색어와 동일/유사어로 검색된 내용이 들어있는 텍스트, 문서, 파일, 데이터베이스 내의 데이터에서 해당문장의 구조를 시멘틱(semantic) 알고리즘을 적용하여 그 의미를 분석한다.

도 3에 본 발명의 두번째와 세번째 단계를 통해 정보를 수집하고 추출하는 과정이 나타나 있다. 도 3에 나타난 바와 같이, 본 발명의 정보 수집, 분석 및 가공 시스템은 웹 문서들을 검색하여 사용자가 입력한 검색어("중국의 인터넷 광고 시장 규모는?")에 해당하는 문서들을 찾아낸 후, 해당 문서로부터 필요한 데이터를 추출한다. 즉, 도 3에 나타난 예에서는 중국의 인터넷 광고시장 규모에 대한 정보가 있는 세 개의 웹 문서를 검색하였으며, 그 문서 중 첫번째 문서에서는 중국의 인터넷 광고 시장규모를 나타내는 표를 추출하고, 두번째와 세번째 문서에서는 기사 중 해당 수치를 나타내는 부분을 추출한다.

이와 같이 추출된 데이터는 데이터 처리부로 전달되어 표준화/정형화되고 데이터베이스에 저장된다(S240). 데이터의 표준화/정형화는 예를 들면 검색된 데이터의 소스에 따라 달리 구성되어 있는 숫자 데이터의 단위를 통일하거나(예를 들면 중국의 화폐 단위인 인민폐(rmb) 단위로 되어 있는 데이터를 불(\$) 단위로 고치는 등), 문자로 표시된 데이터를 숫자로 바꾸는(예를 들면, "이천불" 로 표시된 데이터를 "2,000\$" 로 고치는 등) 등의 작업이 될 수 있다. 또한, 추출된 데이터의 문자, 숫자 등을 분리하여 구조에 따라 데이터베이스를 생성하여 정형화된 형태로 데이터를 저장한다. 이러한 과정의 예가 도 4에 나타나 있다. 도 4에 나타난 바와 같이, 각각 표와 기사의 형태로 나타나 있던 중국의 인터넷 광고시장 규모에 대한 정보가 연도별 시장규모를 나타내는 데이터베이스의 형태로 표준화/정형화되어 데이터베이스에 저장된다.

한편, 사용자가 요구한 정보가 기존에 이미 보관되어 있는 데이터베이스 내에 존재할 수도 있다. 이 경우에는 이 정보를 검색하여 추출한 정보에 덧붙일 수 있다. 도 5는 기존 데이터베이스 내에 미리 저장되어 있는 해외 온라인 광고시장규모 자료를 나타낸다.

다음, 사용자는 자신이 검색하기를 원하는 정보의 표시 형태를 선택한다(S250). 즉, 도 4에 나타난 바와 같은 형태의 매체별/연도별 데이터를 그대로 표로 표시할 수도 있지만, 사용자의 요청에 따라서는 가로축과 세로축을 바꾸어 연도별/매체별의 표로 표시하거나, 매체별로 검색된 시장규모의 평균값을 추가하거나 하는 등으로 표시할 수 있다. 또한, 필요에 따라서는, 데이터베이스에서 데이터를 선택한 후 원하는 표현형태에 맞게 가공(예를 들면, 수식 계산(예를 들면, 월 단위로 표시된 양을 연 단위로 고치는 등))할 수도 있다. 그밖에도, 도 5에 나타나 있는 바와 같은 기존의 데이터베이스를 이용하여 검색을 통해 얻은 정보와 기존 정보를 통합하여 표시하도록 할 수 있다.

또한, 검색된 정보로 웹 화면이나 보고서를 생성할 때, 웹 화면이나 보고서를 어떠한 형태로 꾸밀 것인지를 사용자가 선택할 수 있다. 이러한 과정은 다양한 템플릿을 사용자에게 제공하고 이들 중에서 사용자가 원하는 것을 선택하도록 하거나, 표, 그래프, 스프레드시트 등의 조합 및 배치를 사용자가 직접 디자인하도록 할 수도 있다.

이제, 마지막으로 사용자가 요청하는 형태에 따라 정보를 표시한다(S260).

사용자에게 표시하는 형태는 앞서 S250 단계에서 사용자에게 의해 선택된 대로 웹 화면이나 결과 리포트, 스프레드시트 파일 등 다양한 형태가 될 수 있다. ActiveX 컴포넌트를 이용할 경우, 그리드(grid), 그래프, 보고서, MS-Excel 과 같은 스프레드시트와의 연동이 가능한 형태 등 다양한 표시 형태를 구현할 수 있다.

도 6 내지 도 8은 본 발명의 마지막 단계에서 수집한 정보를 다양한 형태로 사용자에게 표시해주는 것을 나타내는 도면이다. 도 6에서는 정보 수집 결과를 표와 그래프를 포함하는 웹 화면으로 표시해주고 있으며, 도 7은 수집한 정보를 이용하여 자동으로 생성된 전략보고서를 나타내고, 도 8은 수집한 정보를 재처리가능한 스프레드시트 파일(예를 들면 MS Excel)로 포스팅한 것을 나타내고 있다.

보고서를 생성할 때에는, 문장으로부터 원하는 정보를 추출하는 것의 역순으로, 추출되어 데이터베이스에 저장된 정보를 바탕으로 필요한 문장을 생성하여 보고서에 포함되도록 할 수도 있다.

이와 같이 본 발명에서는 단순한 인덱스/링크 정보만을 제공하는 검색 엔진의 경우와 달리 비교/분석/가공된 지식정보를 사용자에게 제시하며, 기존의 기업 MIS, EIS 등의 시스템과 연동하여 전략적 지식 정보를 창출한다.

지금까지 본 발명을 바람직한 실시예를 들어 구체적으로 설명하였으나, 이 실시예는 본 발명을 이해하기 위한 설명을 위해 제시된 것이며, 본 발명의 범위가 이 실시예에 제한되는 것은 아니다. 본 발명의 기술이 속하는 분야의 통상의 전문가라면 본 발명의 기술적 사상의 범위를 벗어나지 않고도 다양한 변형이 가능함을 이해할 수 있을 것이며, 본 발명의 범위는 첨부된 특허청구범위에 의해서 해석되어야 할 것이다.

발명의 효과

이와 같이 본 발명에 따르면, 정보 검색에 드는 시간과 노력을 획기적으로 절약할 수 있으며, 검색 결과의 품질을 향상시킬 수 있고, 또한 기존의 콘텐츠 서비스 업체나, 포털 사이트의 경우 별도의 수작업에 의한 콘텐츠 수집 가공 업무를 줄이고, 서비스 운영 비용을 절약할 수 있다.

또한, 가공된 데이터를 시각적인 형태로 자동으로 보여주므로, 직관적인 의사결정이 가능하고, 리포트 형태의 표시 방법을 이용하면 별도의 보고서 작성 업무를 줄일 수 있다.

(57) 청구의 범위

청구항 1.

컴퓨터 통신망을 통하여 사용자로 부터 검색어를 입력받는 단계,

상기 검색어를 이용하여 상기 검색어가 지시하는 정보가 포함된 웹 문서 또는 데이터베이스를 검색하는 단계,

검색된 상기 웹 문서 또는 데이터베이스로부터 상기 검색어가 지시하는 정보를 추출하는 단계,

상기 추출된 정보를 표준화/정형화하여 데이터베이스의 형태로 저장하는 단계,

데이터베이스의 형태로 저장된 상기 정보를 표시할 형태를 컴퓨터 통신망을 통하여 사용자로부터 입력받는 단계,

데이터베이스의 형태로 저장된 상기 정보를 상기 사용자로부터 입력된 표시형태에 따라 컴퓨터 통신망을 통하여 사용자에게 제공하는 단계를 포함하는 컴퓨터 통신망을 이용한 정보 수집, 가공 및 표시 방법.

청구항 2.

제 1 항에 있어서,

상기 정보를 표시할 형태를 사용자로부터 입력받는 단계에서는,

상기 정보를 포함하는 표, 상기 정보를 나타내는 그래프, 상기 정보를 포함하는 문장, 상기 정보를 포함하는 재가공 가능한 스프레드시트용 파일 중 적어도 하나를 포함하도록 사용자가 표시 형태를 직접 디자인할 수 있는 컴퓨터 통신망을 이용한 정보 수집, 가공 및 표시 방법.

청구항 3.

제 1 항에 있어서,

상기 정보를 표시할 형태를 사용자로부터 입력받는 단계에서는,

상기 정보를 포함하는 다양한 표시 형태를 갖는 템플릿을 컴퓨터 통신망을 통하여 사용자에게 제공하고, 사용자가 상기 템플릿 중에서 원하는 것을 선택하는 컴퓨터 통신망을 이용한 정보 수집, 가공 및 표시 방법.

청구항 4.

컴퓨터 통신망을 통하여 사용자로부터 입력된 검색어를 입력받는 입력부,

상기 입력부로부터 상기 검색어를 제공받아 상기 검색어가 지시하는 정보가 포함된 웹 문서 또는 데이터베이스를 검색하는 데이터 수집 엔진,

상기 데이터 수집 엔진으로부터 검색된 상기 웹 문서 또는 데이터베이스로부터 상기 검색어가 지시하는 정보를 추출하는 정보 추출부,

상기 정보 추출부로부터 추출된 정보를 표준화/정형화하여 저장하는 데이터 처리부,

표준화/정형화된 상기 정보를 저장하는 데이터베이스,

상기 데이터베이스에 저장된 정보를 컴퓨터 통신망을 통하여 사용자에게 제공하는 표시부를 포함하는 컴퓨터 통신망을 이용한 정보 수집, 가공 및 표시 시스템.

청구항 5.

제 4 항에 있어서,

상기 표시부는, 사용자가 상기 정보를 시각적으로 인식할 수 있도록, 상기 정보를 포함하는 표, 상기 정보를 나타내는 그래프, 상기 정보를 포함하는 문장 중 적어도 하나를 포함하는 형태로 상기 정보를 표시하여 제공하는 컴퓨터 통신망을 이용한 정보 수집, 가공 및 표시 시스템.

청구항 6.

제 4 항에 있어서,

상기 표시부는, 상기 정보를 재가공할 수 있는 스프레드시트용 파일로 사용자에게 제공하는 컴퓨터 통신망을 이용한 정보 수집, 가공 및 표시 시스템.

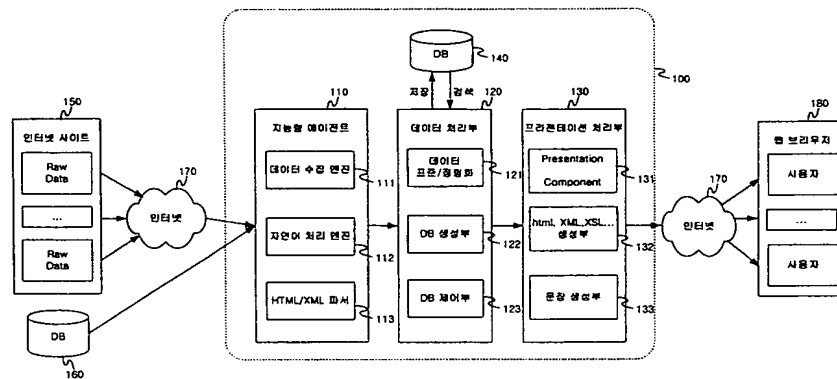
청구항 7.

제 4 항에 있어서,

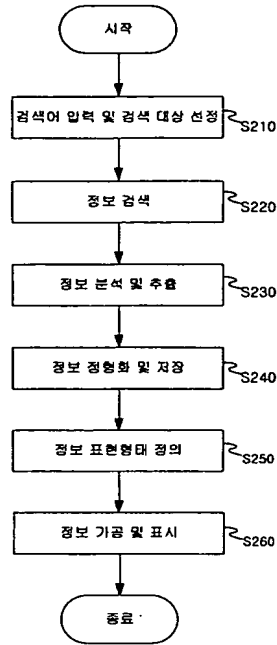
상기 정보 추출부는, 검색된 상기 웹 문서로부터 HTML/XML 태그를 제거하고 정보가 포함된 부분만을 추출할 수 있는 HTML/XML 파서와 검색된 상기 웹 문서에 포함된 문장으로부터 상기 검색어가 지시하는 정보를 추출할 수 있는 자연어 처리부를 포함하는 컴퓨터 통신망을 이용한 정보 수집, 가공 및 표시 시스템.

도면

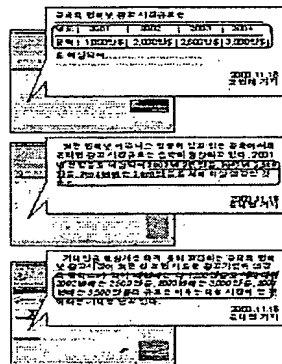
도면 1



도면 2



도면 3



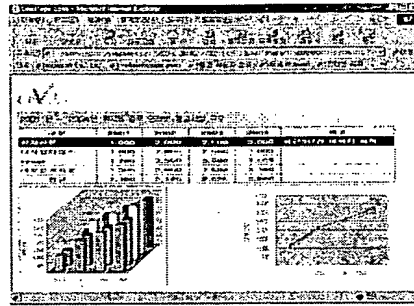
도면 4

2001	2002	2003	2004
1,000	2,000	2,500	4,000
2001	2002	2003	2004
1,000	2,000	2,500	4,000
2001	2002	2003	2004
1,500	2,500	4,000	4,500

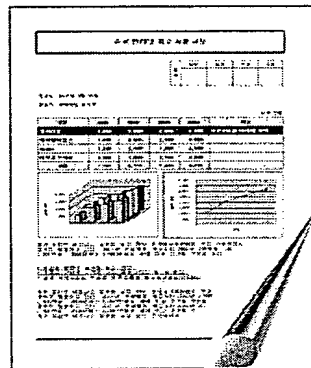
도면 5

구분	2001	2002	2003	2004
비고	9,000	80,000	20,000	80,000
지리	5,000	4,000	6,000	5,000
합계	1,200	2,500	2,000	2,500

도면 6



도면 7



도면 8

구분	2001	2002	2003	2004
전자신문	1,000	2,000	2,500	3,000
디지털타임스	1,000	2,000	2,500	3,000
epage	1,200	2,500	3,000	3,500
내부분석자료	1,200	2,000	2,500	3,300
평균	1,100	2,125	2,625	3,200

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.